

Marquette University

e-Publications@Marquette

College of Communication Faculty Research
and Publications

Communication, College of

7-25-2018

The Ethics of Artificial Intelligence: Superintelligence, Life 3.0 And Robot Rights

Kati Tusinski Berg

Marquette University, kati.berg@marquette.edu

Follow this and additional works at: https://epublications.marquette.edu/comm_fac



Part of the [Communication Commons](#)

Recommended Citation

Berg, Kati Tusinski, "The Ethics of Artificial Intelligence: Superintelligence, Life 3.0 And Robot Rights" (2018). *College of Communication Faculty Research and Publications*. 522.

https://epublications.marquette.edu/comm_fac/522

Marquette University

e-Publications@Marquette

Communication Faculty Research and Publications/College of Communication

This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.

Journal of Media Ethics, Vol. 33, No. 3 (2018): 151-153. [DOI](#). This article is © Routledge (Taylor & Francis Group) and permission has been granted for this version to appear in [e-Publications@Marquette](#). Routledge (Taylor & Francis Group) does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Routledge (Taylor & Francis Group).

The Ethics of Artificial Intelligence: Superintelligence, Life 3.0 And Robot Rights

Kati Tusinski Berg

Department of Strategic Communication, Diederich College of Communication, Marquette University, Milwaukee, WI

In March, I presented on a panel titled "Pressing Issues in Digital Ethics" at the Association for Practical and Professional Ethics 27th Annual International Conference in Chicago. Although the research presented and the discussion generated during this panel had me contemplating many digital ethics concerns, my thoughts kept returning to the ethics of artificial intelligence (AI), particularly given the frightening yet realistic presentation by David Gunkel. Gunkel, who has written extensively on the philosophy of technology, specifically discussed big data, algorithms, and a crisis of responsibility that prompts key questions of values and agency. Since then, I am much more attuned to these issues and notice the topic being discussed more than ever around me. Last week while driving home from campus, I listened to a National Public Radio story titled "Kids, Meet Alexa, Your AI Mary Poppins," about how Amazon worked with child development experts to make Alexa more kid-friendly, which

was a reaction based on consumer interest (Kamenetz, [2018](#)). Earlier this week, in an Issues and Trends email from the Public Relations Society of America, there was a link to a *Fortune* story titled, “4 Big Takeaways from Satya Nadella’s Talk at Microsoft Build.” One of the takeaways was that Microsoft believes in AI and Ethics. Jonathan Vanian ([2018](#)) reported, “Nadella briefly mentioned the company’s internal AI ethics team whose job is to ensure that the company’s foray into cutting-edge techniques like deep learning don’t unintentionally perpetuate societal biases in their products, among other tasks” (para. 11). Advancements in AI will impact all areas of media, including but not limited to journalism, advertising, and public relations. Thus, this trend report focuses on three highly regarded books on AI that are recommended by, and favorites of, both Elon Musk and Bill Gates.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University.

Bostrom’s book has created quite a bit of buzz since it was published in 2014. Elon Musk tweeted about the potential dangers of AI and then penned an open letter expressing his concern to AI researchers, which was signed by Bostrom, Stephen Hawking, and dozens of others. Although Bill Gates did not sign the letter, he has mentioned *Superintelligence* as one of two books we should all read to better understand AI. *Superintelligence* is a serious examination of the risks associated with the likely arrival of machine brains that surpass human brains in general intelligence. In the preface, Brostrom explains, “And, as the fate of the gorillas now depends more on us humans than on the gorillas themselves, so the fate of our species would depend on the actions of the machine superintelligence” (Brostrom, 2014, p. vii). He wanted to explore the challenges of how we might respond to superintelligence because “This is quite possibly the most important and most daunting challenge humanity has ever faced. And—whether we succeed or fail—it is probably the last challenge we will ever face” (Brostrom, 2014, p. vii). Ironically, Brostrom admits that this has not been an easy book to write and, although he hopes it will be an easy book to read, he is not quite sure he succeed.

Brostrom sets the context of his argument by discussing past developments and present capabilities in Chapter 1, yet the bulk of the book studies the kinetics of the an intelligence explosion, the forms and powers of superintelligence, and the strategic choices available to a superintelligent agent that attains a decisive advantage. He then shifts focus to the control problem, reframes the argument to contemplate the larger picture that emerges, and offers some suggestions to increase the chances of avoiding an existential catastrophe later. In a section titled “Seasons of Hope and Despair,” Brostrom provides a historical timeline of AI research that started with the Dartmouth Summer Project in 1956. He concludes that:

The AI pioneers for the most part did not countenance the possibility that their enterprise might involve risk. They gave no lip service—let alone serious thought—to any safety concern or ethical qualm related to the creation of artificial minds and potential computer overlords: a lacuna that astonishes even against the background of the era’s not-so-impressive standards of critical technology assessment. We must hope that by the time the enterprise eventually does become feasible, we will have gained not only the technological proficiency to set off an intelligence explosion but also the higher level of mastery that may be necessary to make the detonation survivable. (Brostrom, 2014, p. 6)

Superintelligence is a short but dense introduction to cutting-edge theory on superintelligence. Some sections were more compelling than others but it is a must-read for those interested in the long-term future of humanity.

Tegmark, M. (2017). *Life 3.0 Being Human in the Age of Artificial Intelligence*. New York, NY: Alfred Knopf.

After 13.8 billion years of cosmic evolution, development has accelerated dramatically here on Earth: Life 1.0 arrived about 4 billion years ago, Life 2.0 (we humans) arrived about a hundred millennia ago, and many AI researchers think that Life 3.0 may arrive during the coming century, perhaps even during our lifetime, spawned by progress in AI. What will happen, and what will this mean for us? That's the topic of this book. (Tegmark, 2017, p. 30)

In his newest book, Max Tegmark, a professor of physics at MIT and president of the Future of Life Institute, writes a compelling guide to what may be the most important conversation of our time: Will AI help us flourish more than ever or give us more power than we can handle? Tegmark's personal storytelling makes the book unintimidating yet powerful, because it demonstrates his commitment to contemplating the both the short-term and long-term ethics and implications of AI. I particularly enjoyed the stories he told to introduce the three distinct schools of thought (digital utopians, technoskeptics, and members of the beneficial-AI movement) and their respective champions (Larry Page, Andrew Ng, and Stuart Russel). In addition to introducing these three distinct perspectives, Tegmark also addresses misconceptions, particularly that the conversation shouldn't be limited to AI researchers. He wrote the book so that everyone will join this conversation and ask:

What sort of future do you want? Should we develop lethal autonomous weapons? What would like to happen with job automation? What career advice would you give today's kids? Do you prefer new jobs replacing the old ones, or a jobless society where everyone enjoys a life of leisure and machine-produced wealth? Further down the road, would you like us to create Life 3.0 and spread it through our cosmos? Will we control intelligent machines or will they control us? Will intelligent machines replace us, coexist with us or merge with us? What will it mean to be human in the age of artificial intelligence? What would you like it to mean, and how can we make the future be that way? (Tegmark, 2017, p. 38)

For those of who are not physicists or AI researchers, Tegmark provides context, visuals and definitions to make the complicated ideas resonate and make sense to readers. Every chapter concludes with a section titled, "The Bottom Line," in which he pulls out key concepts. He also thoughtfully ques his readers as to how it all fits together for both the near future and the aftermath.

As a media ethics scholar, I was particularly intrigued by Chapters 7 and 8, where Tegmark links cold facts to questions of meaning and purpose. He concludes in Chapter 7 that "although many broad ethical principles are agreed upon by most humans, it's unclear how to apply them to other entities, such as non-humans and future AIs" and that it is time "to rekindle research on some of the thorniest issues in philosophy" (Tegmark, 2017, p. 280). And in Chapter 8, Tegmark suggests, "Although we've focused on the future of intelligence in this book, the future of consciousness is even more important, since that's what enables meaning" (Tegmark, 2017, p. 314).

Tegmark provides a unique perspective that doesn't shy away from a full range of viewpoints or from the most controversial issues. The journey he takes readers on in this book is fascinating and I encourage all of you to go along for the ride.

Gunkel, D. (2018). *Robot Rights*. Cambridge, MA: MIT Press.

In his latest book, Gunkel, professor of communication studies at Northern Illinois University, wrote *Robot Rights* "as both a provocation and thoughtful response" (Gunkel, 2018, p. x) to the apparent fact that "the notion of robots having rights is unthinkable" (Levy, [2005](#), p. 393). Gunkel explains the basis of his book in the preface:

It is, therefore, deliberately designed to think the unthinkable by critically considering and making (or venturing to make) a serious philosophical case for the rights of robots. It does so not to be controversial, even if controversy is often the result of this kind of philosophical intervention, but in order to respond to some very real and pressing challenges concerning emerging technology and the current state of and future possibilities for moral reasoning. (p.xi)

Although Brostrom (2014) and Tegmark (2017) allude to ethics in their books by contemplating values, meaning, and purpose; Gunkel confronts the ethics of AI from the beginning of this book and engages a critical examination of philosophical thought throughout the six chapters. He explains that the "majority of work concerning the ethics of artificial intelligence and robots focuses on what philosophers call an agent-oriented problematic;" yet this approach fails to consider the other side of the story (Gunkel, 2018, p. 1). In *Robot Rights*, Gunkel shifts the focus to "consider things from the other side—the side of machine patiency or what Torrance ([2008](#)) called, by contrast, "moral consumers" or the "target" of moral/legal action" (Gunkel, 2018, p. 2). He asks, "Can robots have rights?" and "Should robots have rights?," which raises David Hume's philosophical distinction of the is/ought problem to evaluate and analyze various questions of robot rights. After explaining the critical endeavor he engages in throughout the book, Gunkel asserts,

Developing and debating the rights of robots does not necessarily take anything away from human beings and what (presumably) makes us special; it offers a critical tool for doing work in moral theory, making available new opportunities for us to be more precise and scientific about these distinguishing characteristics and their limits. (Gunkel, 2018, p. 12)

Chapter 1, "Thinking the Unthinkable" enables the reader to get a handle on the terminology, particularly given both the linguistic and conceptual difficulties. Gunkel (2018) asserts:

The terms "robot" and "rights" are already complicated enough, but once you put the two words together, you get something of an allergic reaction. "Robot rights" is for many theorist and practitioners simply unthinkable, meaning that it is either unable to be thought, insofar as the very concept strains against common sense or good scientific reasoning; or is to be purposefully avoided as something that must not be thought—i.e., as a kind of prohibited idea or blasphemy that would open a Pandora's box of problems and therefore must be suppressed or repressed (to use common psychoanalytical terminology) (p. 50).

The bulk of Gunkel's critical inquiry spans Chapters 2–5, in which he gathers "up the best thinking about robots and rights as it is currently formulated by those individuals and groups of individuals who

are best situated to know such things: philosophers, ethicists, jurists, engineers, scientists, etc.” (p. 7). In the final chapter, “Thinking Otherwise,” Gunkel concludes his argument with a “fundamental challenge to ethics and the way moral philosophy has typically defined, decided, and defended the question of the standing and status of others,” which is why I highly recommend *Robot Rights* for any upper division course or graduate seminar on digital ethics (Gunkel, 2018, p. 185). Gunkel’s knowledge of and expertise in philosophy, media studies, technology, and contemporary critical analysis intersect beautifully in his latest book that provides a provocative probe at what was previously considered unthinkable.

References

- Kamenetz, A. (2018, April 30). Kids, Meet Alexa, Your AI Mary Poppins. *National Public Radio*. Retrieved from <https://www.npr.org/sections/ed/2018/04/30/606164343/kids-meet-alexa-your-ai-mary-poppins>
- Levy, D. (2005). *Robots unlimited: Life in a virtual age*. Boca Raton, FL: CRC Press.
- Torrance, S. (2008). Ethics and consciousness in artificial agents. *AI & Society*, 22(4), 495–521. doi:10.1007/s00146-007-0091-8
- Vanian, J. (2018, May 7). 4 big takeaways from microsoft CEO Satya Nadella’s talk at microsoft build. *Fortune*. Retrieved from <http://fortune.com/2018/05/07/microsoft-satya-nadella-build/>